Data Management Planning Part 1: Overview and a USGS Program Experience

Emily Fort
Data and Information Coordinator
USGS National Climate Change & Wildlife Science Center
Reston, VA
February 18, 2015

USGS
science for a changing world

Agenda

1. Overview of data management planning

2. My program's experience

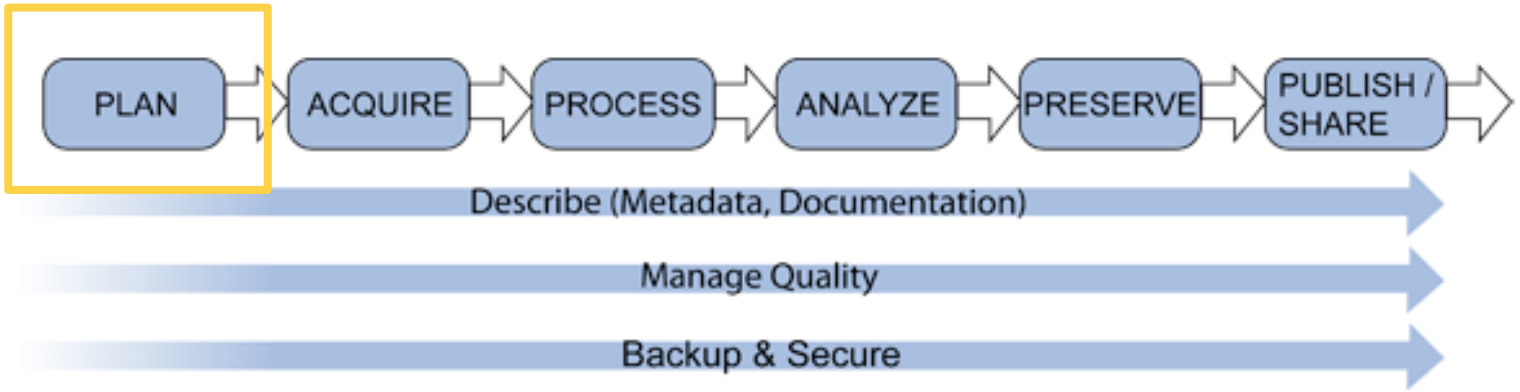
3. Questions!

USGS
science for a changing world

So What is Data Management Planning?

- First step in Data Lifecycle

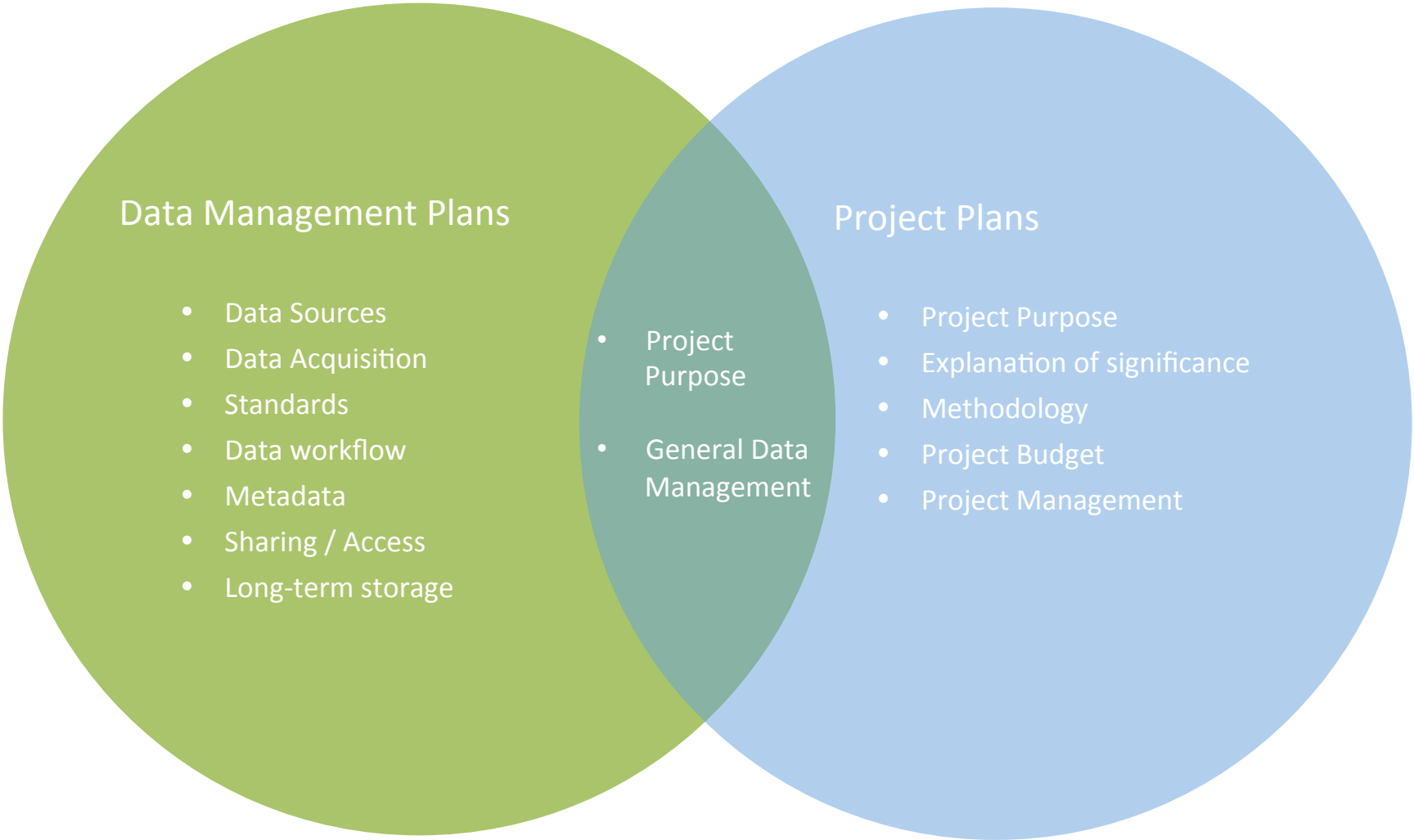
- A data management plan (DMP) is a document that describes what data you will **collect** or **use** and what you will do with your data **during** and **after** your research



Resource: http://www.usgs.gov/datamanagement/plan.php

USGS science for a changing world

How does data management planning fit within project planning?



USGS science for a changing world

Why Should I Care?

- Saves time
- Protects your data investment
  - Ensures that you and others will be able to understand and use data in the future
  - You can get credit when your data are cited
- Increases research efficiency
  - Prevents duplication of effort
- Satisfies funding agency requirements

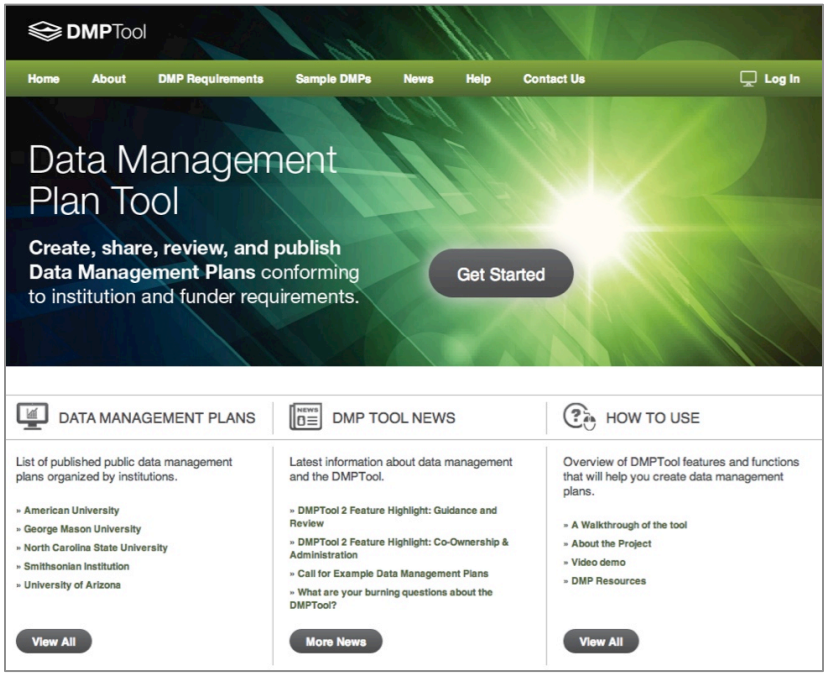
What should be part of my DMP?

- Information about the data
  - Description of data to be produced
  - How will it be managed in short-term?
- Description of Data
  - Format, number of files, approx. volume
    - Processing and quality
- Metadata Content & Format
  - Documentation about the data
- Policies for Access, Sharing, & Reuse
- Long-term Storage & Data Management
  - Where will data be archived?

USGS science for a changing world

How Do I Get Started?

- Many ways to write a DMP
  - One free option is the DMPTool https://dmptool.org/ - USGS template

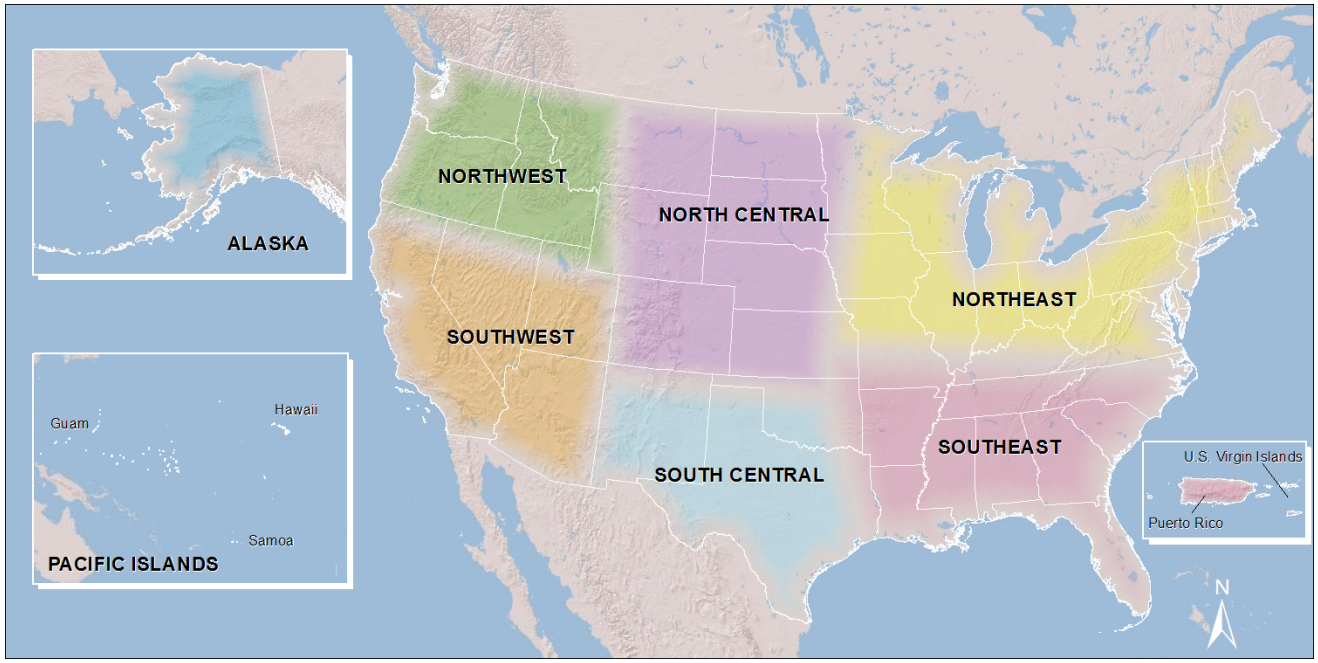


  - NCCWSC has its own template and tool

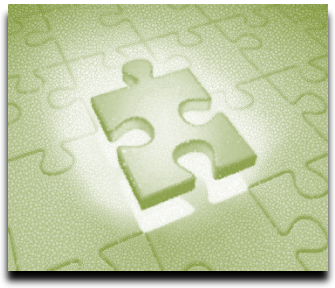
USGS science for a changing world

A Program's Approach

- New program
  - NCCWSC: 2008
  - Climate Science Centers: Beginning in 2010
- Partnerships between USGS and universities
- Fund USGS, other federal, & university scientists



USGS
science for a changing world



Getting Started

- Blank Slate
- Don’t build another stove pipe
- Provide support to CSCs – both tools **and people**
- Develop data management policies
- Link projects to data and products
- Use standards and web services

USGS
science for a changing world

Data Policies

- All project products (data, models, etc.) will be shared (unless there is a good reason not to)
- Sharing happens when the project is complete
- Data management plans are required
- Common standards should be used
- Metadata must be provided
- Developed by group that included federal and university representatives

https://nccwsc.usgs.gov/content/data-policies-and-guidance

USGS science for a changing world

Data Stewards

- Each CSC has a contact who works with funded PIs on their DMPs and the transition of final data products and metadata to a repository
- Meet weekly – status and discussion of various topics (working with tribal data, social science data, sensitive information, etc.)
- Variety of backgrounds – federal, state, and university
- Share experiences and approaches
- Learning as we go!

Data Management Plan Template

- Organized by data inputs (new collections & existing data), models, software/code, and products
- Highlights sensitive issues (sensitive data, data reuse)
- Provides some structure
- Increases awareness of data policies

USGS science for a changing world

Data Inputs

1	[Enter Name of Existing Collection]
Description:	Describe the information that will be used, including its characteristics, temporal scope and scale, and geograph
Source:	Identify the source
Restrictions:	Identify any limitati license restrictions, documentation des
Format:	Identify the format
Fees:	Identify any fees as
Quality Checks:	Identify the proced validation, and an a
Data Processing & Scientific Workflows:	Describe any data p manipulate the dat
Backup & Storage:	Describe the appro project during the p
Volume Estimate:	Estimate the volum
Citation:	Provide citation for

1	[Provide a brief name to describe new data collection]
Description:	Describe the information that will be collected, including its characteristics, temporal scope and scale, and geographic scope and scale, when available.
Data Management Resources:	Describe the proposal resources allocated for data management activities for the new data collected as a level of effort, total dollars allocated, or as a percentage of the total project's cost. Resources could include people's time or proposal funding.
Exclusive Use:	Project data and associated products should be available publicly at the end of the project. If a request to limit access for a period of time after project completion is needed, please identify the length of time and the reason for the extension. (Request cannot be more than one year.)
Restrictions:	Identify any limitations on access or reuse (e.g., sensitive data, restricted data, software with license restrictions, etc.) and provide justification for restriction. Provide citation or documentation describing limitations if due to policies or legal reasons.
Format:	Identify the formats in which the data will be generated, maintained, and made available.
Protocols:	Identify any standard protocols or methodologies that will be used to collect the data, if available.
Quality Checks:	Identify the procedural steps for ensuring data quality.
Data Processing & Scientific Workflows:	Describe data processing steps or provide a scientific workflow you plan to use to manipulate the data, as appropriate.
Metadata:	Identify the metadata standard that will be used to describe the document (FGDC, ISO, EML, etc.)
Volume Estimate:	Estimate the volume of information generated: megabyte (MB), GB, TB, or PB.
Backup & Storage:	Describe the approach for backup and storage of the information associated with the research project during the project.
Repository for Data:	In addition to the NCCWSC repository (ScienceBase), identify any other repositories where you plan to share your data. Indicate if data will be integrated into an existing collection or offered as a new collection.
Citation:	Specify how the project's data should be cited.
Digital Object Identifier (DOI)/Link:	Provide a digital object identifier (DOI)/link to the data when available publicly.
Lifespan of Data	At some point, datasets may be archived. Choose one of the following options to indicate how long you anticipate this data will be of value to other researchers. Less than 5 years,

USGS
science for a changing world

Models & Software

1	[Name of Model]
Description	Provide a brief description of the model and its purpose.
Model Version	Identify the version of model used.
Source/Link:	Provide a link or a source for the model.
Model Input(s)	Enter the types of input data required for driving, calibrating, or validating the model. For proposals, summary information is all that is needed. For funded projects, these should be described in detail in the existing or new collection, data inputs section.
Model Output(s)	Enter the types of output data the model will produce. For proposals, summary information is all that is needed. For funded projects, provide more details as known. If the model output is a generated dataset that is a project deliverable, describe it in detail in the data product section.
Calibration Details	Briefly describe the calibration/validation approach being taken.

1	[Name of Software or Other Need]
Description:	Describe any custom software or code developed or used, and/or any web tools being developed as part of the project.
Source/Link:	If the custom software or code can be accessed via an online repository, provide a link.
Restrictions:	Identify any limitations on access or reuse.
Maintenance and Support for the Web Tool	If a web tool is developed as part of the project, is there a strategy for the ongoing support and maintenance of the web tool after the project is complete? If so, briefly describe it.
Languages:	Identify the computing language/framework that was used (e.g., Java, .Net, Ruby, Rails, SQL, etc.)
Environment:	Identify the operating system environment (e.g., Windows, Linux, MacOS X, etc.)

Data Products

1	[Name of Data Product]
Description:	Describe the information that will be produced, including its characteristics, temporal scope and scale, and geographic scope and scale, when available.
Data Management Resources:	Describe the proposal resources allocated for data management activities for the data products as a level of effort, total dollars allocated, or as a percentage of the total project's cost. Resources could include people's time or proposal funding.
Format:	Identify the formats in which the data will be generated, maintained, and made available.
Exclusive Use:	Project data and associated products should be available publicly at the end of the project. If a request to limit access for a period of time after project completion is needed, please identify the length of time and the reason for the extension. (Request cannot be more than one year.)
Restrictions:	Identify any limitations on access or reuse (e.g., sensitive data, restricted data, software with license restrictions, etc.) and provide justification for restriction. Provide citation or documentation describing limitations if due to policies or legal reasons.
Quality Checks:	Identify the procedural steps for ensuring data quality during the project.
Data Processing & Scientific Workflows:	Describe data processing steps or provide a scientific workflow you plan to use to manipulate the data, as appropriate.
Metadata:	Identify the metadata standard that will be used to describe the data and products (FGDC, ISO, EML, etc.)
Volume Estimate:	Estimate the volume of information generated: megabyte (MB), GB, TB, or PB.
Backup & Storage:	Describe the approach for backup and storage of the information associated with the research project during the project.
Repository for Data:	In addition to the NCCWSC repository (ScienceBase), identify any other repositories where you plan to share your data.
Citation:	Specify how the project's data should be cited.
Digital Object Identifier (DOI)/Link:	Provide a digital object identifier (DOI)/link to the project when available publicly.
Lifespan of Data	At some point, datasets may be archived. Choose one of the following options to indicate how long you anticipate this data will be of value to other researchers. Less than 5 years, 5-10 years, 10-20 years, 20-50 years, 50+ years.

USGS science for a changing world

DMP Review Process

- For each proposal, the proposal's DMP is reviewed by a data steward and comments are provided to the PI
- For funded proposals, an updated and fully completed DMP is required within 1 month of receiving funding
- At project completion, the DMP is reviewed and updated, if necessary

Reviewing DMPs

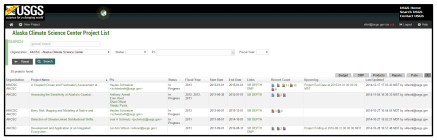
- What do we look for?
  - Is the information in the DMP consistent with the proposed work?
  - Adequate information
  - Is everything complete?
  - Are there any restrictions to data sharing identified?
  - Are resources identified – students? % of budget?
  - Are they using standard formats?
  - Metadata standards?

USGS science for a changing world

Tools to Support Project & Data Management

Project Dashboard (Beta)

Track Project's Status Including Budget, Reports, Publications, Products, & DMP



DMP Editor

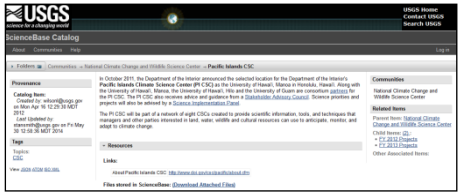
Build/Update DMP

DEPTH

Manage Project Records

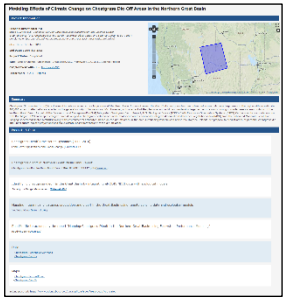
ScienceBase

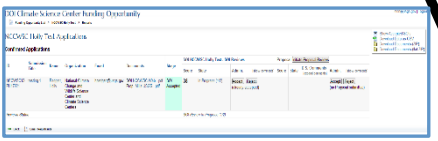
Data Repository for CSC Projects Records and Products



NCCWSC Website

Displays CSC Project Information from ScienceBase using Web Services





RFP Manager

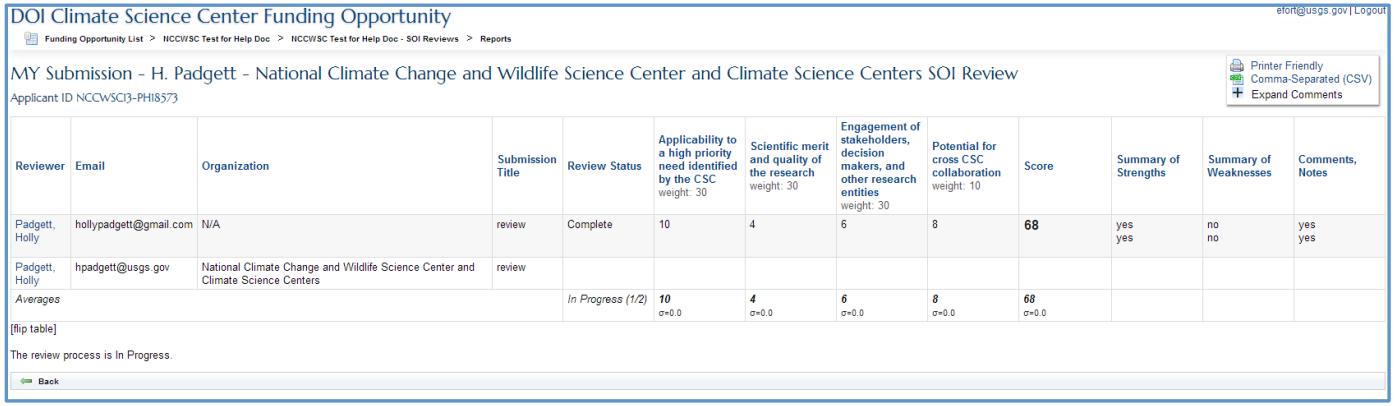
Manages RFP Process and Creates Initial Project Record in Science Base

- Proposal/SOI PDF
- Data Management Plan (DMP)
- Budget Document

USGS science for a changing world

RFP Manager

- Needed a way to collect proposal information and conduct peer review
- Ensures consistency in process and compliance with policies (including DMP)



DOI Climate Science Center Funding Opportunity

efort@usgs.gov | Logout

Funding Opportunity List > NCCWSC Test for Help Doc > NCCWSC Test for Help Doc - SOI Reviews > Reports

MY Submission - H. Padgett - National Climate Change and Wildlife Science Center and Climate Science Centers SOI Review

Applicant ID NCCWSC13-PH18573

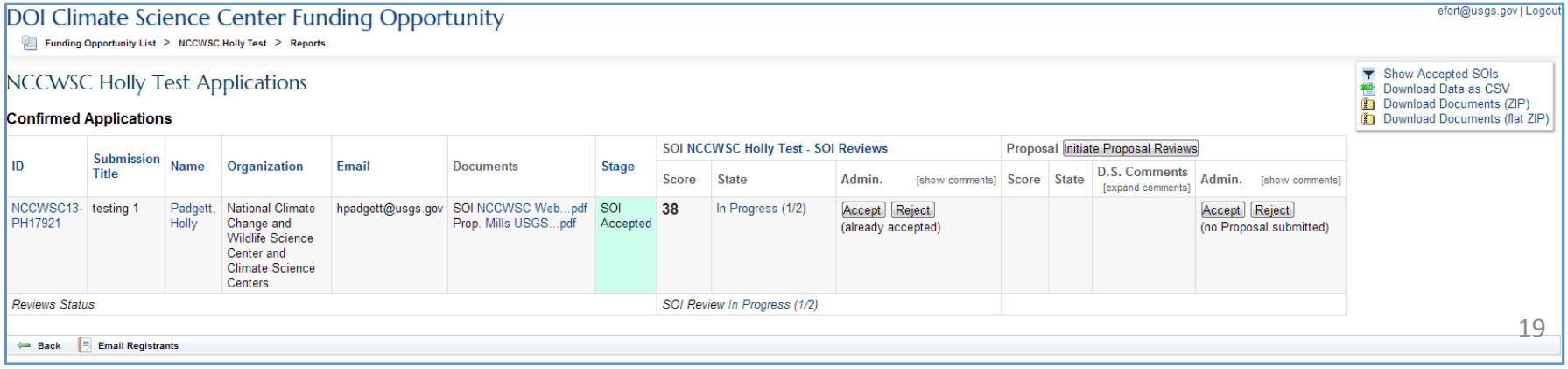
Printer Friendly
Comma-Separated (CSV)
Expand Comments

Reviewer	Email	Organization	Submission Title	Review Status	Applicability to a high priority need identified by the CSC weight: 30	Scientific merit and quality of the research weight: 30	Engagement of stakeholders, decision makers, and other research entities weight: 30	Potential for cross CSC collaboration weight: 10	Score	Summary of Strengths	Summary of Weaknesses	Comments, Notes
Padgett, Holly	hollypadgett@gmail.com	N/A	review	Complete	10	4	6	8	68	yes yes	no no	yes yes
Padgett, Holly	hpadgett@usgs.gov	National Climate Change and Wildlife Science Center and Climate Science Centers	review									
Averages				*In Progress (1/2)*	*10* σ=0.0	*4* σ=0.0	*6* σ=0.0	*8* σ=0.0	*68* σ=0.0			

[flip table]

The review process is In Progress.

Back



DOI Climate Science Center Funding Opportunity

efort@usgs.gov | Logout

Funding Opportunity List > NCCWSC Holly Test > Reports

NCCWSC Holly Test Applications

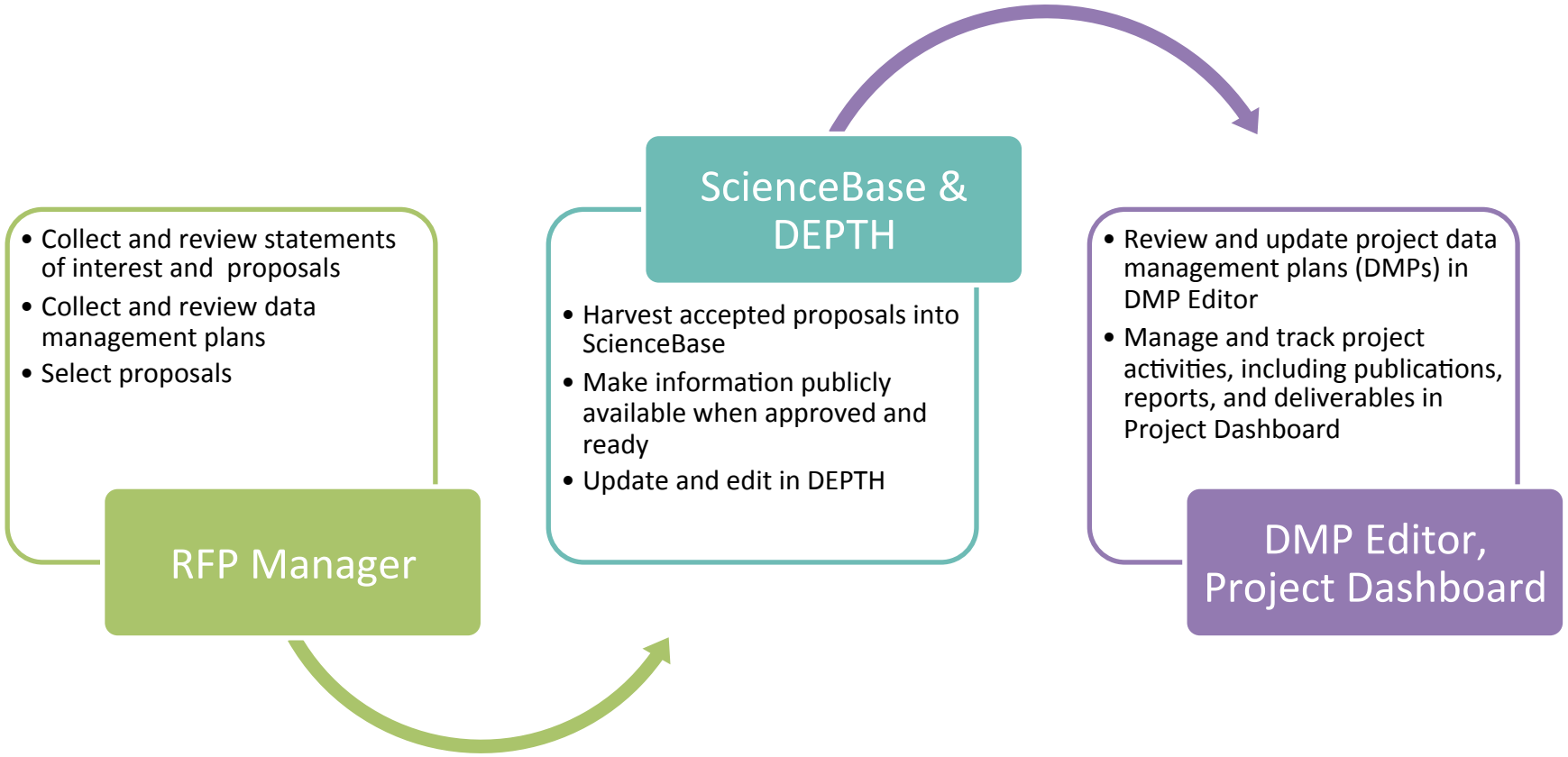
Show Accepted SOIs
Download Data as CSV
Download Documents (ZIP)
Download Documents (flat ZIP)

Confirmed Applications

ID	Submission Title	Name	Organization	Email	Documents	Stage	SOI NCCWSC Holly Test - SOI Reviews			Proposal Initiate Proposal Reviews			
							Score	State	Admin. [show comments]	Score	State	D.S. Comments [expand comments]	Admin. [show comments]
NCCWSC13-PH17921	testing 1	Padgett, Holly	National Climate Change and Wildlife Science Center and Climate Science Centers	hpadgett@usgs.gov	SOI NCCWSC Web...pdf Prop. Mills USGS...pdf	SOI Accepted	38	In Progress (1/2)	Accept Reject (already accepted)				Accept Reject (no Proposal submitted)
Reviews Status							*SOI Review In Progress (1/2)*						

Back Email Registrants

Steps Along the Way



NCCWSC Website Project Pages - Summary

- Organized by CSC and by Fiscal Year Funded
- Icon to indicate map or data/product

Northwest CSC - FY 2011 Projects

Year(s)	Title	Principal Investigator(s)	Contains
2010-2012	Climate Change Threats to Fish Habitat Connectivity: Growth and Predation	Patrick J Connolly (*USGS Columbia River Research Laboratory*)	
2011-2013	Contribution of Landscape Characteristics and Vegetation Shifts from Global Climate Change to Long-Term Viability of Greater Sage-grouse	Steven T Knick (*U.S. Geological Survey, Forest and Rangeland Ecosystem Science Center*) Sara J Oyler-McCance (*U.S. Geological Survey Fort Collins Science Center*)	
2011-2012	Development Support for the NW Climate Science Center Science Agenda	Dar Crammond (*U.S. Geological Survey*)	
2010-2011	Disentangling the Effects of Climate and Landscape Change on Bird Population Trends in the Western U.S. and Canada	Matthew Betts (*Forest Ecosystems and Society, Oregon State University*) Susan Shirley (*Forest Ecosystems and Society, Oregon State University*) Joan Hagar (*U.S. Geological Survey Forest & Rangeland Ecosystem Science Center*)	
2011-2014	Identification and Laboratory Validation of Temperature Tolerance for Macroinvertebrates: Developing Vulnerability Prediction Tools	Robert W Black (*Washington Water Science Center*)	
2011-2012	Modeling Effects of Climate Change on Cheatgrass Die-Off Areas in the Northern Great Basin	Bruce K Wylie (*U.S. Geological Survey Earth Resources Observation and Science Center*) Stephen Boyte (*U.S. Geological Survey, Earth Resources Observation and Science Center & SGT, Inc.*) Donald Major (*Bureau of Land Management Idaho and Great Basin Restoration Initiative*)	
2011-2014	Range-Wide Climate Vulnerability Assessment for Threatened Bull Trout	Jason B Dunham (*U.S. Geological Survey, Forest and Rangeland Ecosystem Science Center*)	

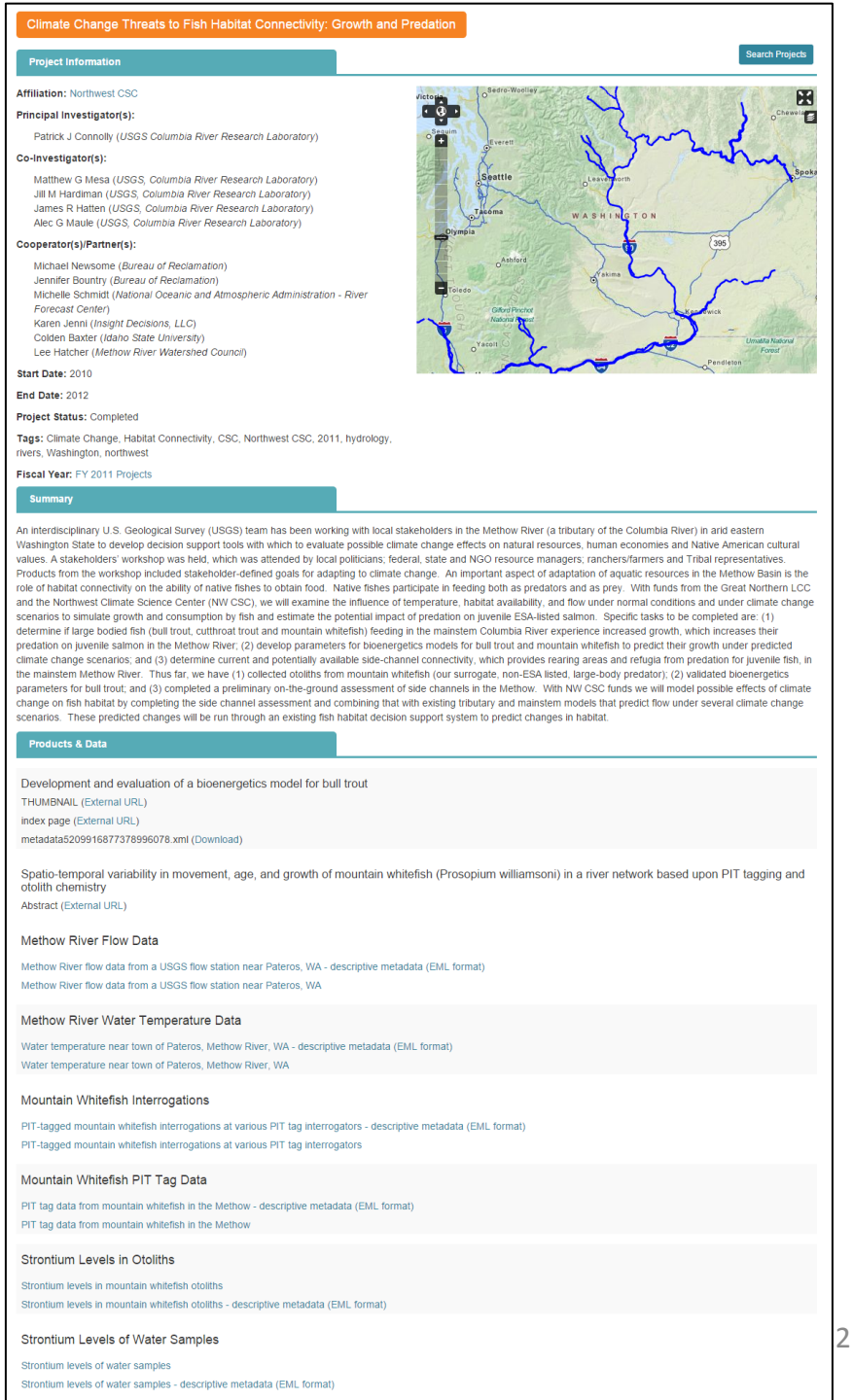
USGS science for a changing world

NCCWSC Website Project Pages - Detail

Powered by ScienceBase

Example: https://nccwsc.usgs.gov/display-project/5006c1e5e4b0abf7ce733f3b/5006f4f4e4b0abf7ce733f96

USGS science for a changing world



Climate Change Threats to Fish Habitat Connectivity: Growth and Predation

Search Projects

Project Information

Affiliation: Northwest CSC

Principal Investigator(s):

Patrick J Connolly (*USGS Columbia River Research Laboratory*)

Co-Investigator(s):

Matthew G Mesa (*USGS, Columbia River Research Laboratory*)
Jill M Hardiman (*USGS, Columbia River Research Laboratory*)
James R Hatten (*USGS, Columbia River Research Laboratory*)
Alec G Maule (*USGS, Columbia River Research Laboratory*)

Cooperator(s)/Partner(s):

Michael Newsome (*Bureau of Reclamation*)
Jennifer Bountry (*Bureau of Reclamation*)
Michelle Schmidt (*National Oceanic and Atmospheric Administration - River Forecast Center*)
Karen Jenni (*Insight Decisions, LLC*)
Colden Baxter (*Idaho State University*)
Lee Hatcher (*Methow River Watershed Council*)

Start Date: 2010

End Date: 2012

Project Status: Completed

Tags: Climate Change, Habitat Connectivity, CSC, Northwest CSC, 2011, hydrology, rivers, Washington, northwest

Fiscal Year: FY 2011 Projects

Summary

An interdisciplinary U.S. Geological Survey (USGS) team has been working with local stakeholders in the Methow River (a tributary of the Columbia River) in arid eastern Washington State to develop decision support tools with which to evaluate possible climate change effects on natural resources, human economies and Native American cultural values. A stakeholders' workshop was held, which was attended by local politicians; federal, state and NGO resource managers; ranchers/farmers and Tribal representatives. Products from the workshop included stakeholder-defined goals for adapting to climate change. An important aspect of adaptation of aquatic resources in the Methow Basin is the role of habitat connectivity on the ability of native fishes to obtain food. Native fishes participate in feeding both as predators and as prey. With funds from the Great Northern LCC and the Northwest Climate Science Center (NW CSC), we will examine the influence of temperature, habitat availability, and flow under normal conditions and under climate change scenarios to simulate growth and consumption by fish and estimate the potential impact of predation on juvenile ESA-listed salmon. Specific tasks to be completed are: (1) determine if large bodied fish (bull trout, cutthroat trout and mountain whitefish) feeding in the mainstem Columbia River experience increased growth, which increases their predation on juvenile salmon in the Methow River; (2) develop parameters for bioenergetics models for bull trout and mountain whitefish to predict their growth under predicted climate change scenarios; and (3) determine current and potentially available side-channel connectivity, which provides rearing areas and refugia from predation for juvenile fish, in the mainstem Methow River. Thus far, we have (1) collected otoliths from mountain whitefish (our surrogate, non-ESA listed, large-body predator); (2) validated bioenergetics parameters for bull trout; and (3) completed a preliminary on-the-ground assessment of side channels in the Methow. With NW CSC funds we will model possible effects of climate change on fish habitat by completing the side channel assessment and combining that with existing tributary and mainstem models that predict flow under several climate change scenarios. These predicted changes will be run through an existing fish habitat decision support system to predict changes in habitat.

Products & Data

Development and evaluation of a bioenergetics model for bull trout
THUMBNAIL (External URL)
index page (External URL)
metadata5209916877378996078.xml (Download)

Spatio-temporal variability in movement, age, and growth of mountain whitefish (Prosopium williamsoni) in a river network based upon PIT tagging and otolith chemistry
Abstract (External URL)

Methow River Flow Data
Methow River flow data from a USGS flow station near Pateros, WA - descriptive metadata (EML format)
Methow River flow data from a USGS flow station near Pateros, WA

Methow River Water Temperature Data
Water temperature near town of Pateros, Methow River, WA - descriptive metadata (EML format)
Water temperature near town of Pateros, Methow River, WA

Mountain Whitefish Interrogations
PIT-tagged mountain whitefish interrogations at various PIT tag interrogators - descriptive metadata (EML format)
PIT-tagged mountain whitefish interrogations at various PIT tag interrogators

Mountain Whitefish PIT Tag Data
PIT tag data from mountain whitefish in the Methow - descriptive metadata (EML format)
PIT tag data from mountain whitefish in the Methow

Strontium Levels in Otoliths
Strontium levels in mountain whitefish otoliths
Strontium levels in mountain whitefish otoliths - descriptive metadata (EML format)

Strontium Levels of Water Samples
Strontium levels of water samples
Strontium levels of water samples - descriptive metadata (EML format)

Where Do We Go From Here?

- Learn lessons from DMPs to improve guidance, clarify, walk line of getting the right amount of information at the proposal stage
- Funding separate projects
  - Social science data management
  - Data from projects involving indigenous peoples
- Add features (DOIs, better metadata tools)
- Analyze DMPs – common datasets, etc

USGS science for a changing world

Questions?

Emily Fort

efort@usgs.gov

703-648-4082

https://nccwsc.usgs.gov/

USGS science for a changing world